Discriminating Between Computer-Generated Facial Images and Natural Ones Using Smoothness Property and Local Entropy

Huy H. Nguyen[1(✉)], Hoang-Quoc Nguyen-Son[2], Thuc D. Nguyen[1], and Isao Echizen[3]

[1] VNUHCM - University of Science, Ho Chi Minh City, Vietnam
honghuy127@gmail.com, ndthuc@fit.hcmus.edu.vn
[2] SOKENDAI (The Graduate University for Advanced Studies), Kanagawa, Japan
nshquoc@nii.ac.jp
[3] National Institute of Informatics, Tokyo, Japan
iechizen@nii.ac.jp

Abstract. Discriminating between computer-generated images and natural ones is a crucial problem in digital image forensics. Facial images belong to a special case of this problem. Advances in technology have made it possible for computers to generate realistic multimedia contents that are very difficult to distinguish from non-computer generated contents. This could lead to undesired applications such as face spoofing to bypass authentication systems and distributing harmful unreal images or videos on social media. We have created a method for identifying computer-generated facial images that works effectively for both frontal and angled images. It can also be applied to extracted video frames. This method is based on smoothness property of the faces presented by edges and human skin's characteristic via local entropy. Experiments demonstrated that performance of the proposed method is better than that of state-of-the-art approaches.

Keywords: Facial image · Computer-generated image · Image forensics · Face spoofing

1 Introduction

Rapid developments in technology have led to major changes in the film and video game industries, particularly in the use of realistic graphics. For instance, it was virtually impossible to distinguish between the real Paul Walker and the computer-generated one in the film "The Fast and the Furious 7"[1]. The death of the actor during filming led the director to use previously recorded digital 3D

[1] http://www.techtimes.com/articles/42216/20150326/hollywood-studios-digitally-scanning-actors-bodies-archival.htm.

Y.-Q. Shi et al. (Eds.): IWDW 2015, LNCS 9569, pp. 39–50, 2016.
DOI: 10.1007/978-3-319-31960-5_4

scan data to reconstruct Mr. Walker's face for the unfinished scenes. Another example is Pro Evolution Soccer[2], a video game developed and published by Konami. Since the 2012 version, the images of the soccer players are rendered so realistically that they look almost like real people.

The identification of computer-generated facial images (and videos) has many applications. Detecting face spoofing is an example. Thanks to morphable model suggested by Blanz and Vetter [1], attackers can now reconstruct 3D images of a person's face from a single 2D frontal image. Unreal images and videos can be used to harm people or to gain political and/or economic advantage. For example, fake images or videos about aliens, disasters, statesmen, or businessmen can create confusion or change peoples' opinions. Social media such as Facebook, Twitter, Flickr, or YouTube is ideal environment to widespread them.

Facial images belong to a special class of images which includes faces of people. Discriminating between computer-generated facial images and natural ones is a specific case of the same problem on general images, which contain any kind of topics such as landscape, architecture, animals, or people. Facial images have some unique attributes of which some approaches for general images do not fully take advantage. These attributes could also degrade the performance of these approaches. In this paper, we focus on facial attributes to maximize the performance.

Our survey about facial images revealed that there are differences in the smoothness property of the faces and skin's characteristic between computer-generated and natural facial images. The smoothness property is reflected in the number of connected components given by an edge detection algorithm follows by morphological closing operation. Natural facial images tend to have more edges which connect to each other, meanwhile edges of computer-generated images are more discrete. Skin's characteristic can also be used in the form of the variation of local entropy. Natural images tend to have smaller variations of local entropy than computer-generated images.

The results of this survey led us to develop a novel method for discriminating between computer-generated facial images and natural ones. It is based on both smoothness property of the faces presented by edges and human skin's characteristic via local entropy. This method works for multi-stage facial images, including frontal and angled ones. For very realistic images, its accuracy is 71.25 %. For well-designed images in a well-known game, its accuracy is 91.23 %. The result is better than that of state-of-the-art methods [3,4,7,8].

The rest of the paper is organized as follow: The related work is introduced in the next section. Continuing, the proposed method are presented with the overview and the two measurements. The experiments and their results are discussed in the evaluation section. The conclusions are drawn in the last section.

[2] https://pes.konami.com/.

2 Related Work

The discriminating between computer-generated images and natural ones topic focuses on two type of images: general images and facial ones. In addition, there are some approaches applying for videos.

2.1 Approaches for General Images

Peng et al. suggested a method for identifying computer-generated images based on the impact of filter array (CFA) interpolation on the photo response non-uniformity noise (PRNU) [10]. The differences of the PRNU correlations between computer-generated images and natural ones are used to discriminating them. The performance is limited by the quality of the noise, which is inferred from various types of filter (Bayer, RGBE, CYYM, etc.).

In other major study, Peng et al. suggested using the colors in images for discrimination [9]. They observed that the colors of natural images are typically more abundant than those of computer-generated ones. The colors are quantified via statistic features (such as histogram or relative frequency) and textural features (e.g. lacunarity, smoothness, entropy, consistency, and multi-fractal dimension). However, this method is not appropriate for facial images because such images have colors that are more balanced.

Lyu and Farid [8] proposed using the statistics of the first and higher-order wavelets. However, wavelet statistics are better suited for natural images than facial images due to the correlation of facial features. Khanna et al. utilized the noise made by digital cameras to efficiently classify not only computer-generated images and camera-produced images but also scanned images [7]. Conotter and Cordin developed a method for measuring the noises using wavelet transformation [3] that works well with both general and facial images. Unfortunately, noise is now being attached to computer-generated images to make them more realistic, and various technologies are now available for removing noise from digital images.

2.2 Approaches for Facial Images

There is only one method for identifying computer-generated facial images proposed by Dang-Nguyen et al. [4]. It is based on the finding that when creating a synthetic face, in most cases, only half of them are made and then duplicated to form a complete one. Post processing may be applied to make it more natural but usually does not change the geometric of the model. Human faces, on the other hand, are not perfectly symmetric. The more symmetric the face is, the high possibility it is generated by computer. The fact remains that symmetric property is detectable in only frontal facial images which limits the scope of this method.

2.3 Approaches for Videos

There are some methods for detecting computer-generated faces in video. In the one developed by Conotter et al. [2], the fluctuations in blood flow are used to distinguish computer-generated from actual faces. However, as the developers pointed out, this physiological signal can be easily simulated by computer to prevent detection. Another method is based on the assumption that facial expressions (such as happiness, sadness, surprise, fear, anger, and disgust) are important factors for recognizing actual faces [5]. A potentially useful characteristic of computer-generated videos is that they often contain repeatable patterns [6]. Unfortunately, these methods work only when there are multiple video frames; they cannot be used for a single image.

3 Proposed Method

3.1 Overview

Our proposed method has three phases, which is illustrated in Fig. 1

Phase 1: *Detect and extract face*
Face is detected and extracted from the input image by using Viola-Jones algorithm. The output is resized to 250 × 250 pixels to ensure that every image is treated equally and to reduce resource consumption. This size sufficiently preserves important patterns of the extracted faces. After that, an ellipse-shaped mask is used to filter out unnecessary parts such as background or hair, which is shown in Fig. 2. The major axis is in vertical direction with 375 pixels in length. The minor axis is in horizontal direction with 200 pixels in length. The intersection of the two axises is at the center of the image.

Phase 2: *Perform measurements*
With the facial image extracted in phase 1, edge-based measurement and entropy-based measurement are performed to obtain data for phase 3. Details about two measurements are presented in the next sections. Edge-base measurement component creates one feature and entropy-based measurement one generates four features for logistic regression.

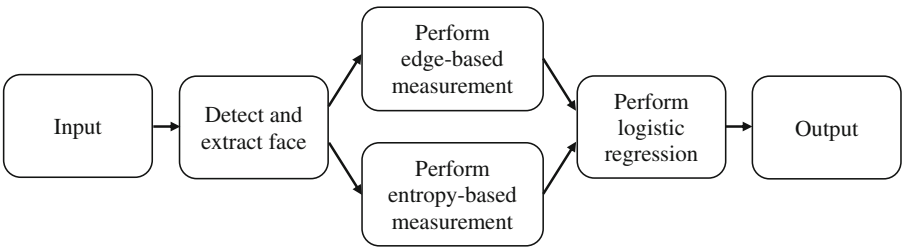

Fig. 1. Overview of proposed method

Phase 3: *Perform logistic regression*
Some well-known machine learning algorithms such as logistic regression, support vector machine (SVM), and sequential minimal optimization (SMO) were evaluated to find the best candidate for the final phase. Logistic regression was chosen because of its best performance. The classification result is "computer-generated image" or "natural image."

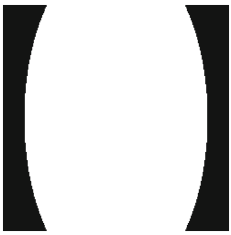

Fig. 2. The ellipse-shaped mask

3.2 Edge-Based Measurement

This phase measures the smoothness of images based on the edge property. Natural images tend to have seamless and smooth connections among facial features and between these features than computer-generated images. This can be measured by the number of connected components obtained by edge detection algorithm. There are three steps in this measurement, which is presented in Fig. 3:

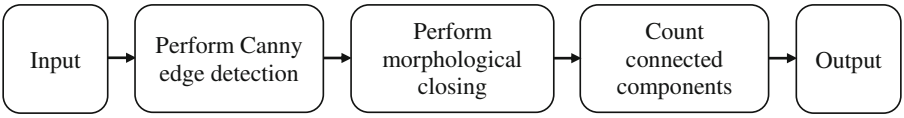

Fig. 3. Overview of edge-based measurement

Step 1: *Perform Canny edge detection*
Canny edge detection algorithm is employed because of its good performance.

The 5×5 Gaussian convolution kernel with $\sigma = 1.3$ is used to remove noise before edge detecting, shown by follows:

$$\frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix}$$

For Sobel operator, we use a pair of 3×3 convolution masks. The first mask G_x estimates the intensity gradients of the image in the horizontal direction (x-direction) and the second mask G_y estimates them in the vertical direction (y-direction), respectively shown as follows:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}; G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

The high and the low threshold used for filtering out some edge pixels caused by noise and color variation are respectively 0.01 and 0.004. We conducted experiment to make sure that these values are optimal, which is presented in evaluation section.

Step 2: *Perform morphological closing*
Morphological closing algorithm is used to fill gaps and to connect related edges together. The structuring element is a disk with 1 pixel radius. This step is significantly important to ensure that related features are connected to each other. Figure 4 illustrates the result of this step.

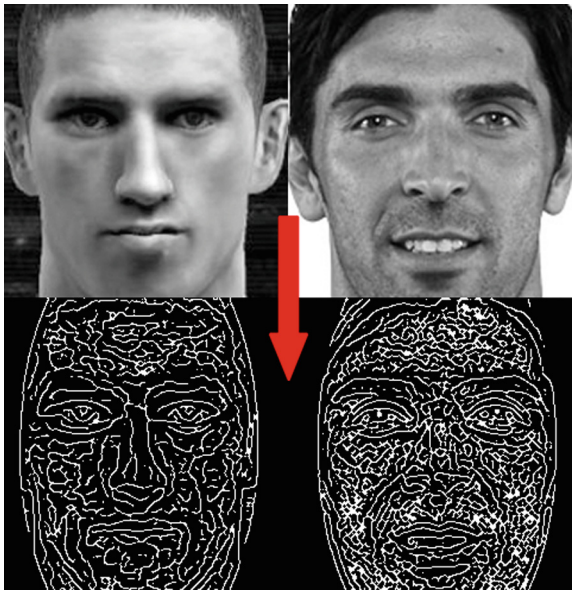

Fig. 4. After perform edge detection and morphological closing algorithm, the natural image on the right have more edges connected together than the computer-generated image on the left

Step 3: *Count connected components*
Connected components are determined and counted with 8-connected neighborhood, illustrated in Fig. 5. A stand-alone is a connected component. A group

0	1	0	0	0	0	0	0	0	3
0	1	0	0	0	0	0	0	3	0
0	0	1	0	0	0	0	0	3	0
0	0	1	0	0	0	0	0	3	0
1	1	1	1	1	0	0	3	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	2	2	2	2
0	0	0	0	1	0	0	0	0	0

Fig. 5. A matrix with three connected components. The component number one has two edges intersect

of edges connecting to each other is also a connected component. Breadth-first search or depth-first search could be employed in this phase.

3.3 Entropy-Based Measurement

This phase measures the variation of local entropy of skin areas in the input images. Based on observations and measurements on facial images, natural ones tend to have smaller variations of local entropy than computer-generated ones. There are three steps in this measurement, which is presented in Fig. 6:

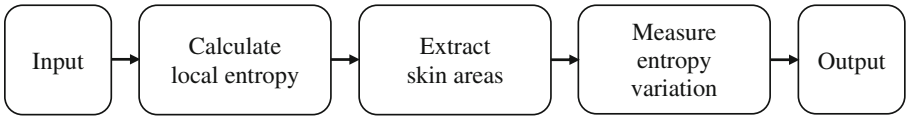

Fig. 6. Overview of entropy-based measurement

Step 1: *Calculate local entropy*
Input images are converted to gray-scale to calculate entropy value of the 9-by-9 neighborhood elements around the corresponding pixel. Symmetric padding is applied for pixels on the borders. Entropy values of the elements of the neighborhood is calculated as:

$$E = -\sum P \circ \log_2(P) \tag{1}$$

where P is the distribution of the elements of the image.

Step 2: *Extract skin areas*
A mask is formed using the entropy matrix by being converted to black and white image with the threshold equal 0.8. Morphological closing algorithm is applied with the 9×9 matrix structuring element, follows by morphological filling holes algorithm. The skin areas of the entropy matrix is extracted by applying this mask. Figure 7 illustrates the result of this step.

Fig. 7. Skin areas extracted from entropy matrix

Step 3: *Measure entropy variation*
A 5×5 window W is moving along the extracted skin image from the left to the right and from the top to the bottom with step S to perform normalization and measurement.

The measurement function is shown by follows:

$$W_{i,j} = \begin{cases} \overline{W} & \text{if } W_{i,j} < \epsilon. \\ W_{i,j}, & \text{otherwise.} \end{cases} \tag{2}$$

where $W_{i,j}$ is the intensity of the pixel at location (i, j), $\overline{W}$ is the average intensity of all pixels in window W, and ϵ is a threshold with small value.

After applying Eq. 2, if the variant of all elements of the window W is less than a threshold T, then W is satisfied the threshold T.

Based on some surveys, we suggest that $S = 2$ and using four couples of ϵ and T:

$$(\epsilon, T) = (2, 2), (2, 4), (2, 8), (5, 5)$$

This step returns the proportion of satisfied windows and total windows. Windows which have all zero pixels are eliminated to improve the accuracy.

4 Evaluation

4.1 Datasets

The datasets were obtained from Dang-Nguyen et al. [4]. There are two datasets of facial images:

– Dataset 1 measures the ability of discriminating between very realistic images and natural images. 40 computer-generated were obtained from the CGSociety website[3] are almost undetectable by human. 40 counterpart natural images were obtained from a variety of sources. Figure 8 shows sample images from dataset 1.

[3] http://www.cgsociety.org/.

Fig. 8. Sample images from dataset 1. Images in top row were computer-generated; those in bottom row are natural

Fig. 9. Sample images from dataset 2. Images in top row were computer-generated; those in bottom row are natural

– Dataset 2 measure the ability of discriminating between computer-generated images rendered in a modern computer game and natural images. It contains 200 computer-generated images from Pro Evolution Soccer 2012[4] and 200 natural images of actual football players. Figure 9 shows sample images from dataset 2.

4.2 Threshold Values Evaluation

We evaluated the proposed edge-based measurement on training data of dataset 1 and dataset 2 with the high thresholds from 0.01 to 0.5. The distance between two adjacent thresholds is 0.05, except for the first one which is 0.04. The low threshold values are 40 % of the high threshold values. The output is classified using logistic regression. The result is illustrated in Fig. 10.

[4] http://www.pesfaces.co.uk.

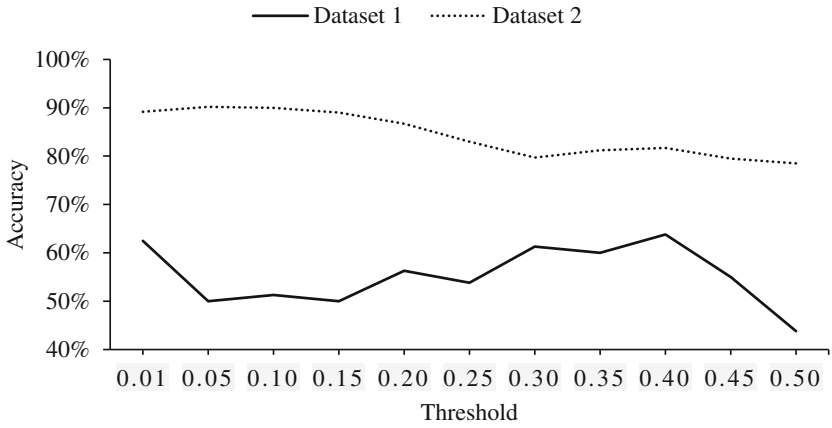

Fig. 10. Accuracy of edge-based measurement on dataset 1 and dataset 2

The high threshold 0.01 is chosen because it gives the best performance on both datasets. The corresponding low threshold is 0.004.

4.3 Experiments

We conducted three experiments: The first experiment is the proposed method with only edge-based measurement; the second is the one with only entropy-based measurement and the third is the full version of the proposed method. The result is shown in Table 1 in comparison with Dang Nguyen's approach [4], the best state-of-the-art one.

Table 1. Classification accuracy on dataset 1 and dataset 2

Approach	Dataset 1	Dataset 2
Dang Nguyen's approach [4]	67.50 %	89.25 %
Edge-based measurement	75.00 %	84.20 %
Entropy-based measurement	62.50 %	89.20 %
Proposed approach	71.75 %	91.23 %

The proposed approach with only edge-based measurement has very good performance on dataset 1 and the one with only entropy-based measurement has better performance on dataset 2. The full proposed approach, which contains the both measurements, has acceptable high performance on dataset 1 with 71.75 % in accuracy and the best performance on dataset 2 with 91.23 % in accuracy. This approach also outperforms Dang Nguyen's [4] and the three other approaches [3,7,8], which is illustrated in Fig. 11.

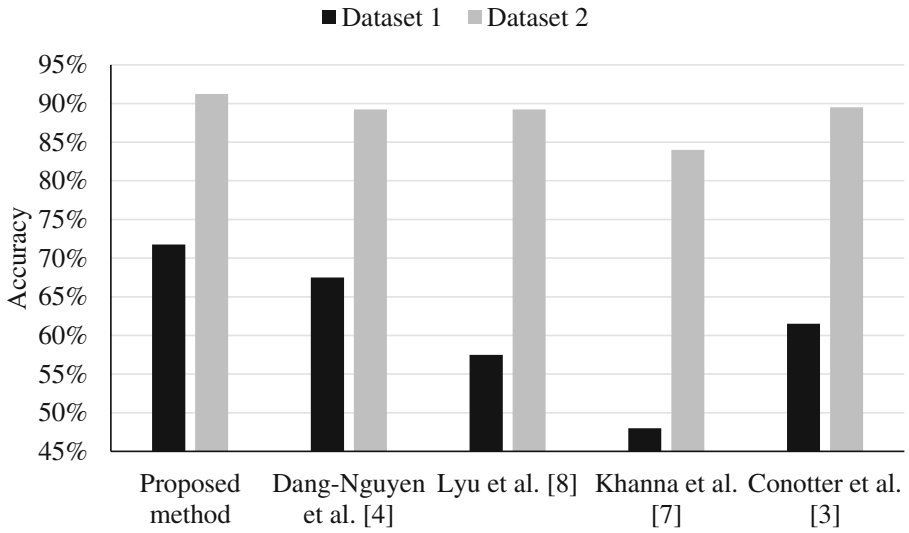

Fig. 11. Comparison of the proposed approach with four other methods using dataset 1 and 2

Images in dataset 1 are well-designed. The characteristic of the skin in this dataset is very similar to human skin. That explains why the performance of entropy-based measurement on dataset 1 is not high. Image resolution also affects edge property. An images with high resolution produces more edges than the same one with lower resolution. Many images in dataset 2, after being extracted faces, has low resolution. The size of the smallest image is 45×45 pixels. The performance of edge-based measurement is significantly influenced by this problem.

5 Conclusion

Our proposed method is an effective way to identify computer-generated facial images. It is based on two properties: the smoothness of the faces presented by edges and the characteristic of human skin via local entropy. Combining the strong points of two measurements, the method performs effectively on multi-stage images and could also be used for extracted video frames. Future work includes optimizing each phase of the proposed method. In particular, the thresholds need to be reevaluated and optimized automatically, and deep learning can be used instead of logistic regression.

Acknowledgments. We would like to thank Dr. Duc-Tien Dang-Nguyen in the Department of Information Engineering and Computer Science (DISI) of the University of Trento, Italy for providing the two datasets used for evaluation.

References

1. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Computer Graphics and Interactive Techniques (SIGGRAPH), pp. 187–194. ACM (1999)
2. Conotter, V., Bodnari, E., Boato, G., Farid, H.: Physiologically-based detection of computer generated faces in video. In: International Conference on Image Processing (ICIP), pp. 248–252. IEEE (2014)
3. Conotter, V., Cordin, L.: Detecting photographic and computer generated composites. In: IS&T/SPIE Electronic Imaging, pp. 7870–7876. SPIE (2011)
4. Dang-Nguyen, D.T., Boato, G., De Natale, F.G.: Discrimination between computer generated and natural human faces based on asymmetry information. In: European Signal Processing Conference (EUSIPCO), pp. 1234–1238. IEEE (2012)
5. Dang-Nguyen, D.T., Boato, G., De Natale, F.G.: Identify computer generated characters by analysing facial expressions variation. In: International Workshop on Information Forensics and Security (WIFS), pp. 252–257. IEEE (2012)
6. Dang-Nguyen, D.T., Boato, G., De Natale, F.G.: Revealing synthetic facial animations of realistic characters. In: International Conference on Image Processing (ICIP), pp. 5327–5331. IEEE (2014)
7. Khanna, N., Chiu, G.C., Allebach, J.P., Delp, E.J.: Forensic techniques for classifying scanner, computer generated and digital camera images. In: Acoustics, Speech and Signal Processing (ICASSP), pp. 1653–1656. IEEE (2008)
8. Lyu, S., Farid, H.: How realistic is photorealistic? IEEE Trans. Signal Process. **53**(2), 845–850 (2005)
9. Peng, F., Li, J.T., Long, M.: Identification of natural images andcomputer-generated graphics based on statistical and textural features. J. Forensic Sci. **60**, 435–443 (2014)
10. Peng, F., Zhou, D.I.: Discriminating natural images and computer generated graphics based on the impact of CFA interpolation on the correlation of PRNU. Digit. Invest. **11**(2), 111–119 (2014)